

Simple and Computerized Discriminant Functions for Difficult Identifications: A Rapid Nonparametric Method

JAMES A. SCOTT AND JON H. SHEPARD

*Natural Resource Ecology Lab., Colorado State University, Fort Collins, Co. 80521
and R. R. #2, Nelson, British Columbia*

Many closely related species cannot be identified using single characteristics alone because some individuals of both species have identical characteristics. Discriminant functions have been developed that use combinations of characteristics to identify species that cannot be identified using single characteristics alone (Dixon, 1965). The procedure consists of measuring many characters for individuals known to belong to each of the two species under study. A discriminant function equation is then found using a computer program. The characteristics of an unknown individual are then measured and inserted into this equation, which provides a number that allows identification when compared to the numbers given by known individuals. The purpose of this paper is to illustrate a rapid nonparametric method and to compare it to traditional linear and nonlinear discriminant functions.

STUDY ANIMALS

Papilio glaucus L. and *P. rutulus* Lucas are known to hybridize in the laboratory and apparently hybridize in British Columbia and South Dakota (Clarke & Sheppard, 1955, 1957, 1962; Brower, 1959a, 1959b). Canadian populations are intermediate in many respects, and *P. glaucus* and *P. rutulus* may in fact be subspecies.

Eight characteristics were quantified for each sex. For males, A, B, C, D, and E (Fig. 1) were measured lengths of male genitalia (all lengths are mm). P describes the form of the prong of Figs. 2-4 (1-no lateral processes; 3-a long lateral process; 2-intermediate). F for males and females is the left forewing length in mm. V for males and females describes the amount of red in the anterior submarginal ventral hindwing light spot (1-all yellow; 2-slightly red; 3-half red; 4-mostly red; 5-all red). For females, G, H, and K are measurements of female genitalic structures (Fig. 7). L describes the form of a lobe (1-leaflike as in Fig. 6; 3-bladelike as in Fig. 5; 2-intermediate). S describes the shade of a flange (Fig. 7) (1-almost transparent; 3-dark gray; 2-intermediate). B describes the dorsal forewing color (1-pale yellow; 3-lemon yellow; 2-intermediate).

A RAPID NONLINEAR NONPARAMETRIC DISCRIMINANT FUNCTION

The simple method involves three steps: 1) choosing characters which differ between the two species. Characters for males were A/B, C/F, E/F, B/C, D/F, P, and V; for females, G/F, L, H/F, S, K/F, V, and B; 2) determining means for each character for each species using individuals known to be correctly identified; 3) characters with means larger in species 1 than in species 2 are multiplied by each other in the numerator of the discriminant function. Characters with means smaller in species 1 than in species 2 are multiplied by each other in the denominator. With *P. rutulus* as species 1 and *P. glaucus* as species 2, the discriminant functions are:

$$\text{SDF } \delta = (B/C \ D/F \ E/F \ C/F) / (A/B \ P \ V),$$

$$\text{SDF } \varphi = (G/F \ K/F) / (LS \ H/F \ VB)$$

COMPUTERIZED DISCRIMINANT FUNCTIONS

These discriminant functions were calculated using the computer program of Dixon (1965). The linear ones are:

$$\text{LDF } \delta = 1.69A - 1.60B - .63C - 2.92D - .23E + .88P \\ - .00056F + .61V$$

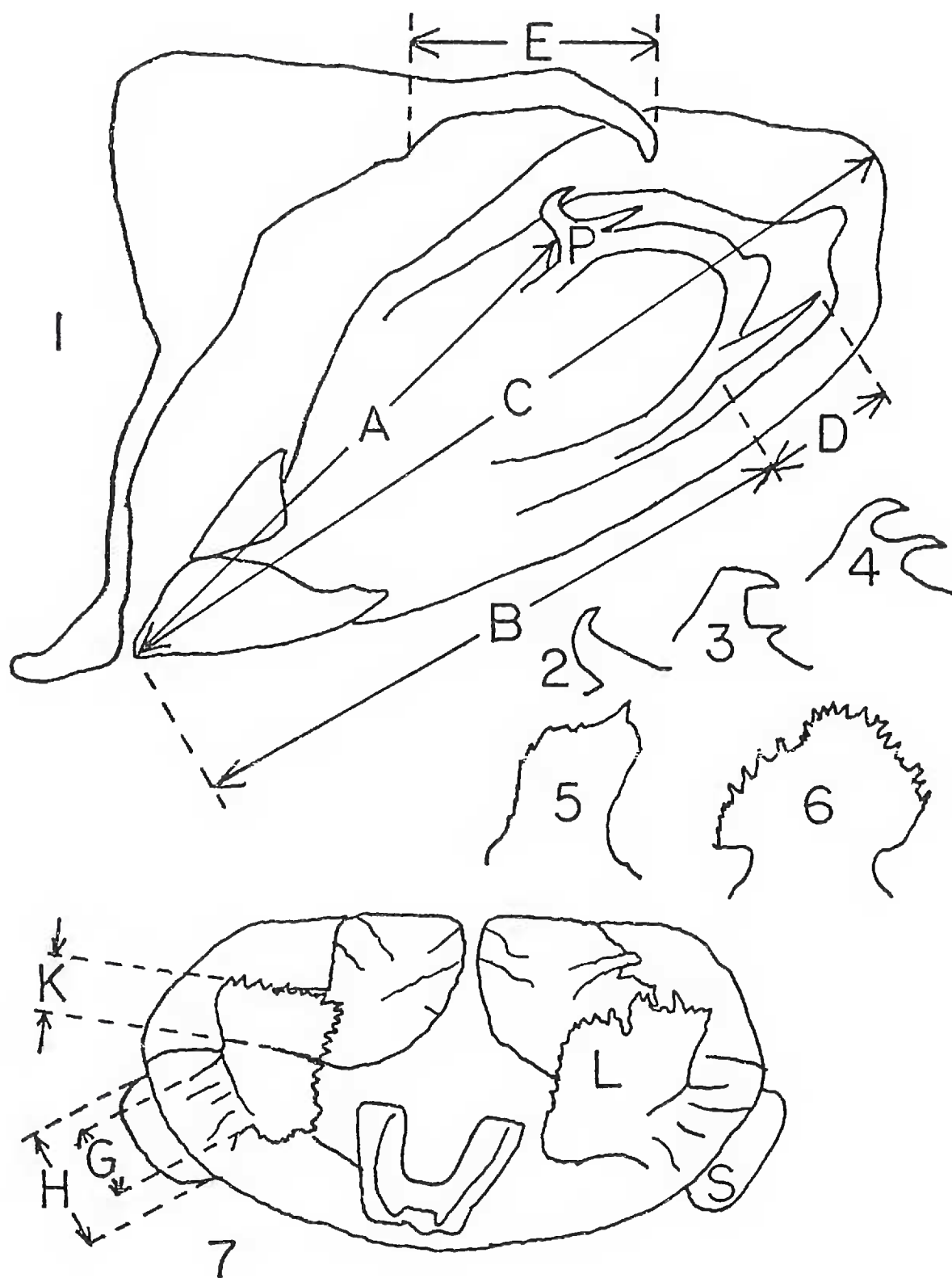
$$\text{LDF } \varphi = .30S - 2.16G + 3.35H + .17K - 1.14F + 1.43V - .22B$$

The nonlinear ones are:

$$\text{NLDF } \delta = (8.34\text{Ln}A - 12.64\text{Ln}B - 6.53\text{Ln}C - 5.13\text{Ln}D \\ - 1.30\text{Ln}E + 3.14\text{Ln}P + 2.75\text{Ln}F + 9.55\text{Ln}V) / 2.30$$

$$\text{NLDF } \varphi = -.71\text{Ln}S + 3.45\text{Ln}G + 3.91\text{Ln}H + .59\text{Ln}K \\ - 21.06\text{Ln}F + 11.47\text{Ln}V + .21\text{Ln}B$$

In all of these discriminant functions to identify an unknown individual the characters (letters) are found, substituted into the equations, and the result is compared to results from individuals of known identity. The results can also be plotted along a single axis for comparison; for the simple function this axis should be logarithmic (Fig. 8). These discriminant functions mathematically maximize the difference between the results for the two species. The computerized functions also minimize the squared deviations from the mean result for the known individuals of each species.



FIGS. 1-7, male and female genitalia. Fig. 1, male genitalia, illustrates measurements A to E, and prong (P). Three shapes of the prong are designated 1 (Fig. 2), 2 (Fig. 3), and 3 (Fig. 4). Fig. 7, female genitalia, illustrates three measurements and the lobe (L) and flange (B). Two shapes of the lobe are designated 1 (Fig. 6) and 3 (Fig. 5).

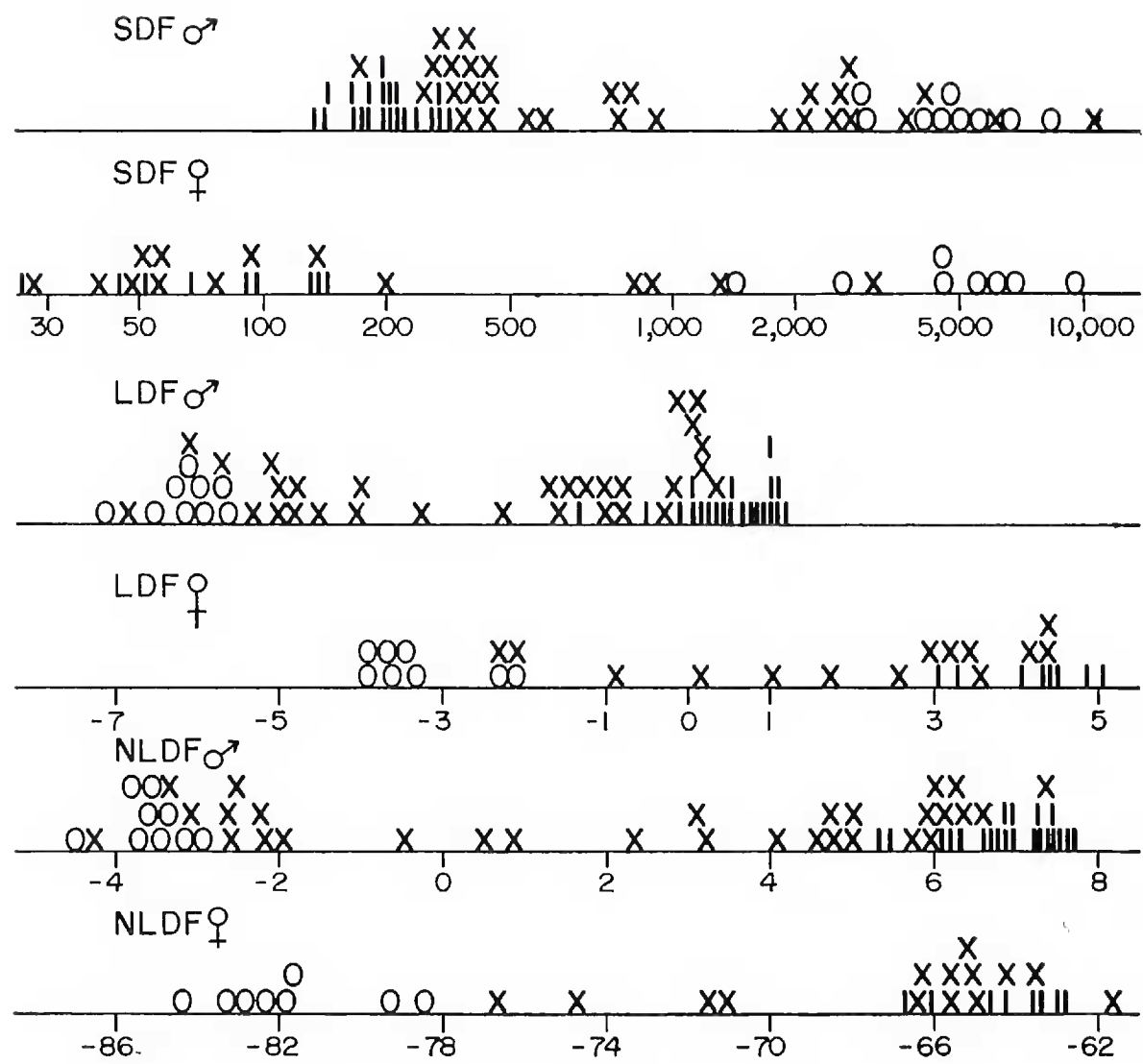


FIG. 8. Discriminant function results for *Papilio glaucus* from eastern United States to Alberta and Alaska (vertical lines), *P. rutulus* from California to Washington (O), and for British Columbia individuals (X). The scale for SDF, and for LDF, is the same for males and females.

ADVANTAGES OF THE RAPID DISCRIMINANT FUNCTION

There are four advantages: 1) it is simple and rapid, without the necessity for computer programs, which may be unavailable; 2) it is nonparametric, which means that the data do not have to conform to a probability distribution as do the computer methods; 3) since the method depends only on the ranking of means, it will not vary with the addition or subtraction of individuals from the group of known individuals from which the means were derived, except in the unlikely event that the ranking of the means of a character for the species changes. If that happens, that character would seem to be of little use for identification, and it should be expunged. The computerized discriminant functions will change somewhat with the addition or subtraction of

each individual, and will stabilize only with large sample sizes. 4) Qualitative (arbitrarily rated numerically) characters are easily used in the simple function, but are very difficult to incorporate into the computerized functions if they do not vary greatly. For example, one of the best characters for females is L, yet its lack of variation in the known individuals prevented it from being used in LDF and NLDF.

Multiplying or dividing any character by a constant (scaling) will not affect the identifications using SDF in any way. The simple discriminant function therefore does not weight characters, whereas the coefficients of the computerized functions are weights. In many taxonomic applications it may be preferable not to weight characters (Sneath & Sokal, 1973), and with small sample sizes of known individuals the computer weights may be unreliable. A possible disadvantage of the simple function is that the computerized methods may provide better identification if large numbers of known individuals are used; this remains to be determined.

RESULTS

All three functions provided excellent identification of the known individuals, and nearly identical identification of the unknowns from British Columbia (Fig. 8). The British Columbia sample includes both "species" and individuals in varying degrees of intermediacy. A further breakdown of the British Columbia sample indicated that central British Columbia individuals were mostly *P. glaucus* and southeastern British Columbia individuals were usually intermediate. We conclude that intergradation does occur between the species in British Columbia. The results are not sufficient to determine whether the intergradation is introgression between species or simply hybridization between subspecies with additive or non-additive (Mendelian) inheritance of characters. The results do provide useful methods for further study of this problem. Full data may be obtained from the authors.

ACKNOWLEDGMENT

The University of California, Davis, provided a small grant for computer time. Specimens are from the collections of J. Shepard, J. Scott, the California Insect Survey, Berkeley, California, and the University of California, Davis, California. Both authors were supported by National Institute of Health traineeships during the initial stage of the study.

LITERATURE CITED

- BROWER, L. P. 1959a. Speciation in butterflies of the *Papilio glaucus* group. I. Morphological relationships and hybridization. *Evolution* 13: 40-63.
- BROWER, L. P. 1959b. Speciation in butterflies of the *Papilio glaucus* group. II. Ecological relationships and interspecific sexual behavior. *Evolution* 13: 212-228.
- CLARKE, C. A. AND SHEPPARD, P. M. 1955. The breeding in captivity of the hybrid *Papilio rutulus* female \times *Papilio glaucus* male. *Lepid. News* 9: 46-48.
- CLARKE, C. A. AND SHEPPARD, P. M. 1957. The breeding in captivity of the hybrid *Papilio glaucus* female \times *Papilio eurymedon* male. *Lepid. News* 11: 201-205.
- CLARKE, C. A. AND SHEPPARD, P. M. 1962. The genetics of the mimetic butterfly *Papilio glaucus*. *Ecology* 43: 159-161.
- DIXON, W. J. [Ed.]. 1973. BMD biomedical computer programs. University of California Press, Berkeley, California. 773 p.
- SNEATH, P. H. A. AND SOKAL, R. R. 1973. Numerical taxonomy. The principles and practice of numerical classification. W. H. Freeman Co., San Francisco, California. 573 p.

ZOOLOGICAL NOMENCLATURE

ANNOUNCEMENT A.N.(S.) 97

Required six months' notice is given of the possible use of plenary powers by the International Commission on Zoological Nomenclature in connection with the following names listed by case number: (see Bull. Zool. Nom. 32, part 3, 22nd September, 1975).

1003. *Chaitophorus* Koch, 1854 (Insecta, Hemiptera): designation of type-species.
2060. *Xiphidium glaberrimum* Burmeister, 1838 and *Orchelimum cuticulare* Audinet-Serville, 1838 (Orthoptera); suppression; designation of *Orchelimum vulgare* Harris, 1841 as type-species of *Orchelimum* Audinet-Serville, 1838.
2107. *Polydrusus* Germar, 1817 (Insecta, Coleoptera): designation of type-species.
2109. *Notozus* Förster, 1853 (Insecta, Hymenoptera, Chrysididae): designation of type-species; *Elampus* Spinola, 1806: proposed suppression.

Comments should be sent in duplicate, citing case number, to the Secretary, International Commission on Zoological Nomenclature, c/o British Museum (Natural History), Cromwell Road, London, S.W.7 5BD, England. Those received early enough will be published in the Bulletin of Zoological Nomenclature. —R. V. MELVILLE, *Secretary to the International Commission on Zoological Nomenclature*.